

Ingenkamp, Karlheinz

**Zur Diskussion über die Leistungen unserer Berufs- und Studienanfänger.  
Eine kritische Bestandsaufnahme der Untersuchungen und Stellungnahmen**

*Zeitschrift für Pädagogik 32 (1986) 1, S. 1-29*



Quellenangabe/ Reference:

Ingenkamp, Karlheinz: Zur Diskussion über die Leistungen unserer Berufs- und Studienanfänger.  
Eine kritische Bestandsaufnahme der Untersuchungen und Stellungnahmen - In: Zeitschrift für  
Pädagogik 32 (1986) 1, S. 1-29 - URN: urn:nbn:de:0111-pedocs-143770 - DOI: 10.25656/01:14377

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-143770>

<https://doi.org/10.25656/01:14377>

in Kooperation mit / in cooperation with:

**BELTZ JUVENTA**

<http://www.juventa.de>

**Nutzungsbedingungen**

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

**Terms of use**

We grant a non-exclusive, non-transferable, individual and limited right to using this document.  
This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

**Kontakt / Contact:**

peDOCS  
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation  
Informationszentrum (IZ) Bildung  
E-Mail: [pedocs@dipf.de](mailto:pedocs@dipf.de)  
Internet: [www.pedocs.de](http://www.pedocs.de)

Digitalisiert

# Zeitschrift für Pädagogik

Jahrgang 32 – Heft 1 – Februar 1986

## I. Thema: Sinkende Schulleistungen – Mythos oder Realität?

- KARLHEINZ INGENKAMP      Zur Diskussion über die Leistungen unserer Berufs- und Studienanfänger. Eine kritische Bestandsaufnahme der Untersuchungen und Stellungnahmen      1
- WALTER H. SCHREIBER      Methoden und Ergebnisse überregionaler Lernerfolgskontrollen in westlichen Industrieländern      31
- HORST DICHANZ/  
RON PODESCHI      Krise im amerikanischen Schulwesen?      51

## II. Thema: Geschichte der Berufsbildung

- JÜRGEN SCHRIEWER      Intermediäre Instanzen, Selbstverwaltung und berufliche Ausbildungsstrukturen im historischen Vergleich      69
- KLAUS HARNEY/  
HEINZ-ELMAR TENORTH      Berufsbildung und industrielles Ausbildungsverhältnis – Zur Genese, Formalisierung und Pädagogisierung beruflicher Ausbildung in Preußen bis 1914      91

## III. Diskussion

- ANDREAS KNAPP      Der Zusammenhang von elterlichem Engagement für Schule und Schulleistung mit Einstellungen und Selbstwahrnehmungen der Kinder      115

## IV. Besprechungen

ANDREAS FLITNER

JÜRGEN HABERMAS: Der philosophische Diskurs der Moderne. Zwölf Vorlesungen

JÜRGEN HABERMAS: Die neue Unübersichtlichkeit. Kleine politische Schriften V 129

FRANZ-MICHAEL  
KONRAD

HELMUT FEND: Die Pädagogik des Neokonservatismus 134

JÜRGEN OELKERS

SIEGFRIED MÜLLER/HANS-UWE OTTO: Verstehen oder Kolonialisieren? Grundprobleme sozialpädagogischen Handelns und Forschens 139

ULRICH HERRMANN

JOACHIM RITTER/KARLFRIED GRÜNDER (Hrsg.): Historisches Wörterbuch der Philosophie 142

INGRID LOHMANN

WOLFGANG KLAFKI: Neue Studien zur Bildungstheorie und Didaktik. Beiträge zur kritisch-konstruktiven Didaktik 145

## V. Dokumentation

Pädagogische Neuerscheinungen 151

# Contents

## I. Topic: Downward Achievement Trends

- KARLHEINZ INGENKAMP      The Discussion about the Achievement of Students Entering Vocational Training or Universities – A Critical Review of Research and Statements in the Federal Republic of Germany      1
- WALTER H. SCHREIBER      On the Methods and Results of Large-Scale Assessment Programs of Student Achievement in Western Industrial Countries      31
- HORST DICHANZ/  
RON PODESCHI      Crisis in American Education?      51

## II. Topic: History of Vocational Education

- JÜRGEN SCHRIEWER      Intermediate Bodies, Self-Government, and Patterns of Vocational Education: a Comparative Historical Analysis of 19<sup>th</sup> Century France and Germany      69
- KLAUS HARNEY/  
HEINZ-ELMAR TENORTH      Profession-forming and industrial vocational education relations – on the genesis and pedagogical transformation of vocational training in Prussia up to 1914      91

## III. Discussion

- ANDREAS KNAPP      The relation between parental involvement in school achievement and their children's self-assessment and attitudes towards school      115

## IV. Book Reviews      129

## V. Documentation

- New Books      151

# Zur Diskussion über die Leistungen unserer Berufs- und Studienanfänger<sup>1</sup>

*Eine kritische Bestandsaufnahme der Untersuchungen und Stellungnahmen*

## *Zusammenfassung*

Der Beitrag gibt einen methodenkritischen Überblick über die in der Bundesrepublik veröffentlichten Untersuchungen zum Leistungsstand unserer Berufs- und Studienanfänger sowie zu Leistungsveränderungen zwischen verschiedenen Jahrgängen. Durch einen historischen Verweis auf die über Jahrhunderte belegten Klagen über den Leistungsabfall wird zunächst auf einen einseitigen Beobachtungsfehler, einen Verklärungseffekt, hingewiesen.

Im krassen Gegensatz zur pädagogischen und gesellschaftlichen Bedeutung, die Lern- und Erziehungserfolge unseres Bildungssystems haben und die sich in der öffentlichen Beachtung dieses Themas niederschlägt, ergibt die Bestandsaufnahme keine systematischen und den gegenwärtigen methodischen Möglichkeiten entsprechenden Erfolgsstudien. Die wenigen, methodisch meist unzureichenden Untersuchungen können die häufig geäußerten Klagen über den Leistungsabfall weder stützen noch widerlegen.

## *1. Einleitung*

Auch 1984 gab es für die Medien wieder einen Anlaß, ein altes Dauerthema zu beleben: die Klagen über die nachlassenden Leistungen unserer Jugend. Jetzt standen die *Studienanfänger* im Mittelpunkt, und den Anlaß boten die Umfrage des Hochschulverbandes (HELDMANN 1984) und das Tutzingener Referat des WRK-Präsidenten (BERCHEM 1984). BERCHEM faßte die Klagen der Universitäten zusammen: „Studienanfänger seien in ihrer Mehrzahl nicht fähig, schwierigen Lernsituationen standzuhalten, sich zu konzentrieren, sie seien wenig belastbar, sie besäßen Kommunikationsschwierigkeiten und verfügten kaum über Lerntechniken“ (1984, S. 6). Das ähnelt sehr der Feststellung von 1837: „Daher zeigen auch unsere Gymnasialschüler bei dargebotener Gelegenheit, daß sie nur das Erlernte wissen, daß ihre Denkkraft zu wenig geübt ist, daß es ihnen an nichts mehr als eigenen Gedanken fehle“ (zitiert nach PAULSEN 1896/97, S. 373). Wer ähnliche Stellungnahmen aus den letzten zwei Jahrhunderten sucht, findet sie in genügender Zahl bei ALBERT 1953, FURCK 1975 oder INGENKAMP 1967<sup>b</sup>.

Bei den *Berufsanfängern* ist die Beurteilung ähnlich negativ, aber konkreter. „In Deutsch und Rechnen mangelhaft“ (*Kammerreport* 1981, S. 153) ist meist der Kern der Aussage. Man ist sicher: „Der Kenntnisstand der Ausbildungsbewerber in gerade diesen Kulturtechniken hat sich im Laufe der letzten Jahre ganz erheblich verschlechtert“ (ohne Autor in: Betriebliche Ausbildungspraxis 1982, S. 929). Auch hier ließe sich die Reihe der Zitate weit in die Vergangenheit verlängern (vgl. ARNHOLD/SENFT 1930; GRÜNER 1984; LESCHINSKY 1982 u. STEMME 1954).

In ihrer Gesamtheit widerlegen sich diese Aussagen selbst. Wenn dieser seit Jahrhunderten beklagte Leistungsabfall tatsächlich so stattgefunden hätte,

dann müßte nach dem Maßstab von 1820 die heutige Professoren- und Ausbildungsgeneration unfähig sein, einfache Sätze niederzuschreiben oder einfache Maschinen in ihren Konstruktionsprinzipien zu verstehen. Die außerordentlichen Fortschritte in Wissenschaft und Technik – und auf die Voraussetzungen dafür bezogen sich die Klagen – beweisen, daß die Aussagen so nicht mit der Realität übereinstimmen können. In diesen Klagen scheint sich eine Art psychischer Selbstschutz des alternden Menschen auszuwirken, der die (eigene) Vergangenheit verklärt und die Gegenwart und Zukunft zunehmend kritischer und skeptischer sieht. Dieser *Verklärungseffekt* täuscht auch unsere Erinnerung über die eigenen Schulleistungen. Ungewollt addieren wir ihnen späteren Wissenserwerb hinzu. Selten sieht die ältere Generation mit genügend Selbstkritik, daß sie *anderes wußte* und *anderes nicht wußte*, daß sie andere Untugenden und andere Tugenden hatte. In der unbewußten Rivalität zur Jugend ist es leichter, die differenzierte Andersartigkeit auf das einfache „Weniger“ in Bereichen zu reduzieren, denen wir uns gewachsen fühlen. Dieser Verklärungseffekt – hier nur mit wenigen Strichen skizziert – ist nur allzu menschlich. Aber er darf uns den Blick nicht trüben, wenn es um nachprüfbare Aussagen über die Leistungsveränderung der Jugend geht. Die Frage ist für unsere Gesellschaft zu wichtig, und sie ist auch mit dem Hinweis auf Verklärungseffekte noch nicht beantwortet. Selbst wenn ein kontinuierliches Absinken der Leistungen (und Fähigkeiten) über Jahrhunderte hinweg durch die Fakten widerlegt wird, so könnte es vorübergehende Abfälle und Anstiege gegeben haben, und wir könnten gegenwärtig in einer abfallenden Phase sein. Da uns unser subjektiver Eindruck so leicht in die Irre führt, müssen wir um so sorgfältiger prüfen, was an wissenschaftlichen Beweisen für eine Veränderung der Leistungen und der Leistungsfähigkeit vorgelegt wird.

## 2. Methodische Anforderungen an Leistungsvergleiche

In der zu referierenden Literatur werden die Maßstäbe der empirisch orientierten Sozialwissenschaften nicht abgelehnt, sondern für den eigenen Bereich behauptet oder zumindest gefordert. Man will mit repräsentativen Untersuchungen „die Voraussetzungen für eine sachliche, an Fakten orientierte und dadurch vorurteilslose Diskussion schaffen“ (DIHT 1967, S. 4), man will empirisch darstellen und vergleichen, ja sogar damit „Grundlagenforschung“ leisten (IW 1981 b, S. 4). Aber im Gegensatz zu diesen Äußerungen werden methodische Standards in kaum einer Untersuchung beachtet.

Aus Raumgründen können die methodischen Anforderungen hier nicht dargestellt werden. Sie sind in allen einschlägigen Lehrbüchern beschrieben (z. B. BORTZ 1984; INGENKAMP 1985; SCHULTZ u. a. 1981). In dem hier zu referierenden Bereich sind es vor allem Methoden

- zur Sicherung der Repräsentanz von Stichproben,
- zur Erfüllung der Gütekriterien von Untersuchungsinstrumenten, insbesondere Objektivität, Zuverlässigkeit und Gültigkeit,
- zur Vergleichbarkeit der Untersuchungsbedingungen,

- zur statistischen Absicherung der Unterschiede zwischen verschiedenen Ergebnissen, insbesondere bei Veränderungsmessungen.

Methodische Standards sind kein Selbstzweck, sondern sie wurden entwickelt, um unseren Erkenntnisgewinn nicht durch vermeidbare Irrtümer zu belasten. Der Nachweis, daß die Standards erfüllt wurden, ist eine *Bringeschuld* des untersuchenden Wissenschaftlers. Die Aussage: „Eine repräsentative Untersuchung, die voll den jeweils anerkannten Standards empirischer Sozialforschung genügt, ist eine Utopie“ (MÖNIKES 1981, S.289f.) ist eine reine Schutzbehauptung. Solche Utopien sind in anderen Untersuchungsbereichen oder in anderen Ländern (vgl. SCHREIBER 1986 in diesem Heft) oft genug realisiert worden.

### 3. Untersuchungen an Berufsanfängern

#### 3.1. Querschnittuntersuchungen

Ein wichtiges Motiv für die zahlreichen Untersuchungen an Berufsanfängern war die Abwehr von Vorwürfen gegen die Ausbildungsbetriebe. Man wollte zeigen, daß die Berufsausbildung nicht mehr leisten kann, wenn die Schule bei der Vermittlung von „Kulturtechniken“ versagt. Die meisten Untersuchungen sind einmalige *Querschnittstudien*, auch wenn sie Aussagen über Leistungsveränderungen enthalten. Sie sind bisher nicht methodenkritisch zusammengefaßt worden, denn BUNK (1980, o. S.) will in einem Überblick nicht in eine „Reliabilitäts- und Validitätsprüfung“ eintreten, und in einer anderen Zusammenfassung stellt er die Frage nach wissenschaftlicher Kritik an den Schluß (BUNK 1983), ohne eine Antwort zu versuchen. Aus Raumgründen werden die Querschnittuntersuchungen tabellarisch dargestellt (siehe *Tab. 1*).

Tab. 1: Querschnitt-Untersuchungen zum Leistungsstand der Berufsanfänger

Autor/ Jahr	Region	Stichprobe	n	Angabe zur Repräsen- tanz	Untersuchungs- bereiche	Verfahren	Angabe zu Gütekrite- rien	Ergebnisse	Prüfstata- tistik
Walther 1929	Rheinpro- vinz	Volksschul- abgänger	150	K. A.	Rechnen Rechtschreiben	4 Rechenaufgaben 8 Sätze Diktat	K. A.	Rechnen: Jungen 56% Mädchen 50% richtig Rechtschreiben: Jungen 7,8 Fehler Mädchen 6,7 Fehler	K. A.
Heilandt 1929	15 Werk- schulen, vor allem in Preußen	Lehrlinge 3 Mon. nach Schulabgang	611	K. A.	Rechnen Rechtschreiben	11 Rechenaufgaben Diktat	K. A.	16,8% mangelhaft u. ungenügend im Rechnen, 37,6% im Diktat	K. A.
IHK Bonn 1953	Kammerbe- reich Bonn	Lehr- abschluß- prüfungen	795	K. A.	Abschlußprüf. Rechnen, Aufs., Buchführ., Schriftverkehr	K. A.	K. A.	Berufsspezifisch. Insgesamt unbefriedigend	K. A.
Stemme 1954	NRW	Berufs- anfänger	20133	K. A.	Rechnen Rechtschreiben	7 Rechenaufg. Diktat	K. A.	Rechnen: 19,1% O.F., 8,7% 6 + 7 F., Diktat: 8,4% 0 + 1 Fehler, 33,9% 11 + mehr Fehler	K. A.
DIHT 1966, 1967	Bundesgebiet	Volksschul- abgänger mit Abschluß	2134	Lehrlinge mit Anfangsbuch- st. G., sonst vage Angaben	Rechnen Rechtschreiben	7 Rechenaufg. (I), 4 eingekl. Rechenaufg. (II), Diktat	K. A.	Rechnen I 22-59% Versager, Rechnen II 26,1% nicht ausr. Diktat 20,5% mangelhaft	K. A.
Glatz 1967	1. Soltau 2. Hannover	Volksschüler 9. Klasse	1: ca. 200 2: 416	K. A.	Rechnen Rechtschreiben	wie DIHT 1966	K. A.	Nur 3,6% bzw. 4,3% mangelhaft in Rechnen, 0 bzw. 1,2% mangelhaft im Diktat	K. A.



Tab. 1 – Fortsetzung

Autor/ Jahr	Region	Stichprobe	n	Angabe zur Repräsen- tanz	Untersuchungs- bereiche	Verfahren	Angabe zu Gütekrite- rien	Ergebnisse	Prüfstä- tistik
Queisner 1967	IHK Pfalz	Lehrfinge (Quartaner als Kontroll- gruppe)	582 (+ 155)	K. A.	Rechnen Rechtschreiben	Wahrscheinlich RT 8 + u. RST 8 + aus der Reihe „Deutsche Schultests“	Aus der Lit. bekannt	AM = 37 Punkte Rechnen (Eichstichprobe = 30,3 Punkte) AM = 36 Fehler Rechtschreiben (Eichstichprobe = 33,3 Fehler)	K. A.
IHK Essen u. a. 1968	Essen, Mül- heim (Ruhr), Oberhausen	Lehrfinge	5020	K. A.	Rechnen Rechtschreiben	7 Rechenaufgaben Diktat	K. A.	Je 44% mangelhaft in Rechnen und Rechtschreiben	K. A.
Brand u. Maisch 1973	IHK Aachen	Berufs- anfänger	536	K. A.	Rechnen Rechtschreiben	Berufsanfängertest	K. A.	Gutes Gesamtergebnis, differenziert nach Berufsgruppen	K. A.
IHK Stade 1975	Kammerbe- zirk Stade	Berufs- anfänger	550	K. A.	Rechnen Rechtschreiben	Grundwissentest	K. A.	19,27% nicht ausreichend in Rechn., 1,64% in Deutsch	K. A.
o. Autor 1977, ident. m. Kurhess. Wirtsch. 1976	Mehrere IHK'n in Hessen	Ausbildungs- anfänger	4364	K. A.	Rechnen Deutsch	Rechentest 40 Aufg., Rechtschreibtest 80 Aufgaben	K. A.	19,6% nicht ausreichend in Rechnen, 17,1% in Deutsch	K. A.
Freytag 1978	IHK'n in Hessen	Ausbildungs- anfänger	4851	K. A.	Rechnen Deutsch	Rechentest 36 Aufg., Rechtschreibtest 80 Aufgaben	K. A.	15,7% nicht ausreichend in Rechnen, 22,1% in Deutsch	K. A.
Freytag 1979	IHK'n in Hessen	Ausbildungs- anfänger	7421	K. A.	Rechnen Deutsch	Rechentest 36 Aufg., Rechtschreibtest 80 Aufgaben	K. A.	28,2% nicht ausreichend in Rechnen, 15,4% in Deutsch	K. A.

Tab. 1 – Fortsetzung

Autor/ Jahr	Region	Stichprobe	n	Angabe zur Repräsen- tanz	Untersuchungs- bereiche	Verfahren	Angaben zu Gütekrite- rien	Ergebnisse	Prüfstä- tistik
o. Autor 1980 (Kurhess. Wirtsch.)	IHK'n in Hessen	Ausbildungs- anfänger	6209	K. A.	Rechnen Rechtschreiben	Rechentest 36 Aufg., Rechtschreibtest 80 Aufgaben	K. A.	24,4% nicht ausreichend in Rechnen, 10,9% in Rechtschreiben	K. A.
o. Autor 1981 (Kurhess. Wirtsch.)	IHK'n in Hessen	Ausbildungs- anfänger	6639	K. A.	Rechnen Rechtschreiben	Rechentest 36 Aufg., Rechtschreibtest 80 Aufgaben	K. A.	20% nicht ausreichend in Rechnen, 24,2% in Rechtschreiben	K. A.
Freytag 1982	IHK'n in Hessen	Ausbildungs- anfänger	5033	K. A.	Rechnen Rechtschreiben	Rechentest 36 Aufg., Rechtschreibtest 80 Aufgaben	K. A.	21,7% nicht ausreichend in Rechnen, 29,3% in Rechtschreiben	K. A.
Freytag 1983	5 IHK'n in Hessen	Ausbildungs- anfänger	4481	K. A.	Rechnen Rechtschreiben	Wie bei Freytag 1979	K. A.	26,6% nicht ausreichend in Rechnen, 15,5% in Rechtschreiben	K. A.
IHK Hanau 1980	Kammerbe- reich Hanau	Lehranfänger	ca. 1200	K. A.	Rechnen Rechtschreiben	Rechentest 36 Aufg., Rechtschreibtest 80 Aufgaben	K. A.	Fast 25% nicht ausreichend in Rech- nen, etwa 12% in Rechtschreiben	K. A.
Metzger 1980	IHK Mittl- er Neckar	Ausbildungs- anfänger	658 Rech- nen 324 Recht- schr.	K. A.	Rechnen Rechtschreiben	7 Zahlenaufg. 5 Textaufg. 2 Geometrieaufg. Lückentest 120 Sätze	K. A.	22,5% nicht ausreichend in Rechnen, 13,2% in Rechtschreiben	K. A.
Schöne- berg 1981	1 Gewerb- liche Berufs- schule	Berufschul- anfänger	438	K. A.	Rechnen	15 Rechenaufgaben	K. A.	13–61% richtige Lösungen je nach Schulabschluß	K. A.

Tab. 1 – Fortsetzung

Autor/ Jahr	Region	Stichprobe	n	Angabe zur Repräsen- tanz	Untersuchungs- bereiche	Verfahren	Angabe zu Gütekrite- rien	Ergebnisse	Prüfstata- tistik
o. Autor 1981 (Un- sere Wirt- schaft)	5 IHK'n in NRW	Ausbildungs- anfänger	5726	K. A.	Rechnen Rechtschreiben	20 Rechenaufgaben Diktat	K. A.	40% mangelhafte Lösungen in Rech- nen u. 44% im Diktat (nach Presse- dienst der IHK'n)	K. A.
o. Autor 1983 (Wirtsch. u. Berufs- Erz.)	5 IHK'n in NRW	Ausbildungs- anfänger	4900	K. A.	Rechnen Rechtschreiben	K. A.	K. A.	Fast ein Drittel in Rechnen mangel- haft, 57% im Diktat	K. A.
Göbel 1982 u. IW 1981	Bundesgebiet	Auszu- bildende	4796	Geschichtete Quotenstich- probe	Rechnen, Rechtschr. (Ar- beitslegenden, Werthaltungen)	34 Rechenaufgaben 24 Aufgaben aus dem Mannheimer Rechtschreibtest	K. A.	Im Rechnen etwa 71,2% der Aufg. richtig, keine mittl. Angabe für Recht- schreiben	z. T.

K. A. = Keine Angabe

Die tabellarische Zusammenstellung erhebt nicht den Anspruch auf Vollständigkeit, da Untersuchungen dieser Art oft in grauer Literatur oder in nur regional verbreiteten Zeitschriften erscheinen. Es fehlt aber keine in dieser Literatur als wichtig angesehene Untersuchung.

In allen registrierten Untersuchungen werden methodische Standards kaum beachtet. Es ist die ganz seltene Ausnahme, wenn sich Angaben finden, die die wissenschaftliche Seriosität der Untersuchung belegen und eine Überprüfung ermöglichen. Ausländische Untersuchungen und neuere Forschungsmethodik scheinen weitgehend unbekannt zu sein. Es ist sicher ernst gemeint, wenn FREYTAG (1979, S. 43) fragt, „ob es denn wirklich ein so ‚wissenschaftliches‘ Problem ist, festzustellen, ob ein Prüfling die Aufgabe ‚6397:89‘ zu lösen imstande ist“. Wenn FREYTAG das allein wissen wollte, brauchte er nur zu prüfen, ob seine Aufgabenstellung (mit Vorgabe von Antworten) durch vergrößerte Ratechancen usw. zu anderen Lösungshäufigkeiten führt als das eigene Ausrechnen, die Freie-Antwort-Form. Aber FREYTAG will ja gar nicht feststellen, ob *ein* Prüfling diese *eine* Aufgabe löst. Er will von dieser einen Aufgabe auf das für die Berufsausbildung wichtige „Rechnen“ *insgesamt* schließen und sieht seinen einen Prüfling als *repräsentativ* für viele an. Er kommt durch Vergleich mit nach Augenschein ähnlichen Aufgaben zu Aussagen über Leistungsveränderungen und folgert auf Gründe, die in Lehrplänen, didaktischen Methoden und der Schulorganisation liegen. Ob derartige Schlußfolgerungen unbewiesenes Gerede oder hinreichend gesicherte Beweise sind, hängt von den Methoden ab, mit denen Aufgaben- und Personenstichproben ausgewählt, untersucht und ausgewertet werden.

In forschungsmethodischer Hinsicht sind hier – anders als in den meisten Bereichen – keine Fortschritte über die Jahrzehnte hin zu erkennen. Im Gegenteil, frühe Arbeiten sind vorsichtiger in ihren Interpretationen. HEILANDT (1929, S. 100) erörtert noch Maßstabprobleme, bei den neueren Autoren taucht nie ein Hinweis auf Probleme der Bestimmung von Kriteriumsgrenzen auf (vgl. HERBIG 1976; KLAUER 1983); nie wird gesehen, daß Aufgaben auch durch Lehrer in ihrem Schwierigkeitsgrad nur ungenau einzuschätzen sind (vgl. SCHNOTZ 1971; vgl. für die Unterschätzung des tatsächlichen Leistungsniveaus eindrucksvoll HOPF 1980). In den letzten drei Jahrzehnten hat sich jedoch die Bezeichnung für Rechenaufgaben und Diktate geändert. Sie werden jetzt „Test“ genannt, obwohl sie nicht einmal die Meßqualität informeller Tests (INGENKAMP 1985) nachweisen. Die EDV, auf die oft hingewiesen wird, hat dazu geführt, daß Prozentverteilungen für alles und jedes berechnet werden, was Vergleiche eher erschwert und nachträgliche statistische Analysen oft unmöglich macht. Der Ausgawert dieser Untersuchungen ist nach wissenschaftlichen Kriterien minimal. Es soll daher nur kurz auf drei Arbeiten ein wenig näher eingegangen werden, um dieses Urteil zu belegen. Die 1966 vorgelegte Untersuchung des DIHT war der aner kennenswerte erste Versuch, repräsentative Daten für das Bundesgebiet zu gewinnen.

Die Versuche, die Stichprobenrepräsentanz zu sichern, waren jedoch unzureichend. Die Untersuchungsinstrumente wurden nicht überprüft, die Verarbeitung, Prüfung und Interpretation der Daten entsprach auch den damaligen Standards in keiner Weise, wie ein-

gehend nachgewiesen worden ist (INGENKAMP 1967a). Die Einwände gegen die curriculare Gültigkeit wurden häufig vorgebracht (vgl. zum Beispiel GLATZ 1967 und WITTMANN 1967). Bedrückend ist, daß aus diesen Fehlern nichts gelernt wurde.

Einen neueren großen Untersuchungskomplex bilden die Untersuchungen einiger Industrie- und Handelskammern in Hessen (FREYTAG und ohne Autor in: Kurhessische Wirtschaft).

Der erste Einwand richtet sich gegen die Stichprobenbildung, die – wie man 1981 erfährt – 12 Prozent aller Auszubildenden in Hessen erfaßt, aber als anfallende Stichprobe ohne Beweis und gegen jede Wahrscheinlichkeit als repräsentativ angesehen wird. Der zweite Einwand betrifft die Untersuchungsinstrumente, die im Rechnen und beim Korrekturlesen in Antwort-Auswahl-Form vorgegeben werden, ohne daß je etwas über Vorerprobungen und Gütekriterien gesagt wird. Der Rechtschreib-„Test“ enthielt Sätze wie diesen „Der Hase (1) *erschauerde*<sup>A</sup> = *erschauerte*<sup>B</sup> = *erschaurte*<sup>C</sup> im Hochholz beim (2) *Anblick*<sup>A</sup> = *anblick*<sup>B</sup> = *Anblik*<sup>C</sup> = *Anblik*<sup>D</sup> = *Anblig*<sup>E</sup> des (3) *rießenhaften*<sup>A</sup> = *riesenhaf-ten*<sup>B</sup> = *rißenhaften*<sup>C</sup> = *riesenhaf-ten*<sup>D</sup> Wesens,“ ... usw. mit weiteren fünf Aufgaben in einem Satz. BUNK (1980, o. S.) hat nach einer Wortschatzüberprüfung festgestellt, daß dieses Verfahren „zum großen Teil aus einem Ratespiel“ besteht. Wenn es auch bei entsprechender Vorerprobung berechtigt ist, Rechtschreibkenntnisse in der Form des Korrekturlesens zu überprüfen, so wissen wir doch seit langem, daß die Vorgabe von Alternativangeboten den Schwierigkeitsgrad erheblich verändern kann (vgl. INGENKAMP 1962). Das wurde in den Berichten nie reflektiert. Bis 1979 wurde von dieser „Test“-Leistung auf „Deutsch“ geschlossen. Ein dritter Einwand muß sich gegen die nie begründete Zuordnung von Noten zu Fehlern und damit gegen die Versagensgrenze richten. Wenn man schon die Problematik norm- versus kriterienorientierter Messung außer acht läßt, so müßte man wissen, daß bei Anlehnung an eine Normalverteilung der Ergebnisse 25 Prozent die Note 5 und 6 erhalten (Stasix). Daß das pädagogisch nicht motivierend wäre, steht auf einem anderen Blatt. Aber warum dann Prozentanteile um 20 Prozent nicht ausreichender Leistungen so bestürzt aufgenommen werden, bleibt unverständlich.

Die Querschnitte in Hessen erfolgten bis 1980 mit jährlich *anderen* Aufgaben an *unterschiedlichen* Stichproben. Ohne jede methodische Absicherung werden dennoch Veränderungen interpretiert (Vierter Einwand). In der *Kurhessischen Wirtschaft* (ohne Autor 1981, S. 143) liest man dann überrascht, daß 1980 zum erstenmal ein Vergleich über einen Zeitraum von drei Jahren möglich war und entnimmt FREYTAG (1983, S. 14), daß ein „früher – nämlich 1978 – benutzter Test wiederverwendet (wurde), so daß die Ergebnisse voll vergleichbar sind“.

Eine Zusammenstellung der Prozentanteile nicht ausreichender Leistungen ergibt folgendes Bild:

Ein eindeutiger Trend ist nicht festzustellen. Da Ergebnissprünge wie zum Beispiel in Rechnen von 1978 zu 1979 und in Rechtschreibung von 1980 bis 1981 und von 1982 auf 1983 als gültiges Indiz echter Trends in der Schule unwahrscheinlich sind, können die Un-

Tab. 2: Prozentanteile nicht ausreichender Leistungen in den hessischen Untersuchungen (Angabe nach Berichtsjahren)

	1977	1978	1979	1980	1981	1982	1983
Rechnen	19,6	15,7	28,2	24,4	20	21,7	26,6
Rechtschreiben	17,1	22,1	15,4	10,9	24,2	29,3	15,5

terschiede vermutlich durch Verschiedenheit der Stichprobe und/oder der Instrumente erklärt werden. Die Interpretation, so ist schließlich einzuwenden, erörtert das nicht, sondern greift mit Vorliebe auf die negativen Wirkungen schulischer Reformen zurück.

Die Untersuchung des Instituts der Deutschen Wirtschaft (GÖBEL 1982; IW 1981 a – c), die vom BMW veranlaßt und finanziert wurde, wird in dieser Literatur häufig als methodisch anspruchsvoll und sozialwissenschaftlichen Forschungsstandards entsprechend bezeichnet. Tatsache ist, daß hier der Stichprobenziehung besondere Aufmerksamkeit gewidmet wurde. Keinerlei Fortschritte sind dagegen bei den Untersuchungsinstrumenten, den Untersuchungsbedingungen und der Interpretation in den publizierten Berichten erkennbar.

Der Rechen-„Test“ bestand aus 34 Aufgaben mit je fünf Auswahlantworten. In dem publizierten Ergebnisbericht (GÖBEL 1982) wird behauptet, sie „beruhen auf einer Analyse der Lehrpläne der Klasse 7–9 der Haupt- und Realschulen“ (S. 40), und die Übereinstimmung von Aufgaben und Lehrplänen sei durch Lehrer beurteilt. Keinerlei Daten über die Erprobung der Aufgaben, über Zuverlässigkeit und Gültigkeit werden mitgeteilt, und auch Hinweise auf Informationen, denen mehr zu entnehmen wäre, fehlen. Sogar im umfangreichen Abschlußbericht (IW o. J.) fehlen Daten zu Gütekriterien.

Der Rechtschreib-„Test“ bestand aus einer Zufallsauswahl von je vier Aufgaben aus den sechs Subtests des Mannheimer Rechtschreibtests (JÄGER/JUNDT 1973, <sup>2</sup>1981). Auch hier werden keine Gütekriterien für diese Kurzauswahl von 24 Aufgaben mitgeteilt. Die nachträgliche Berechnung mit Hilfe der Angaben im Beiheft des M-R-T ergibt für die 24 Aufgaben einen Schwierigkeitsgrad (Md) von 71 und eine Reliabilität von 0,65. Der hohe Meßfehler aufgrund dieser geringen Reliabilität wird in den Interpretationen nicht berücksichtigt.

Das INSTITUT DER DEUTSCHEN WIRTSCHAFT (IW) wollte auch „Einstellungen und Verhaltensweisen“ untersuchen, beschränkt sich aber auf „Werthaltungen“ und „Arbeitstugenden“. Zur Erfassung der Arbeitstugenden wurden zwölf Substantive wie „Zielstrebigkeit“, „Initiative“, „Disziplin“ vorgegeben, bei denen auf einer siebenstufigen Skala (keine bis sehr hohe Bedeutung) angekreuzt werden mußte, welche Bedeutung sie für eine erfolgreiche Ausbildung haben.

Zur Erfassung der Werthaltungen „Extraversion – Introversion“, „Altruismus – Egoismus“ usw., wurden in einem „Sprechblasentest“ acht gezeichnete Situationen vorgelegt, auf die durch Auswahl einer von vier vorgegebenen Feststellungen reagiert werden mußte (GÖBEL 1982, S. 179). Da Leistungen im Mittelpunkt dieses Beitrages stehen, kann zu diesen Skalen und ihrer Verwendung für viele Analysen ohne Angabe irgendwelcher Gütekriterien nur gesagt werden, daß so etwas wissenschaftlichen Standards nicht entspricht und in keiner seriösen fach-psychologischen Zeitschrift so publiziert werden könnte.

Zu den *Durchführungsbedingungen* geben GÖBEL (1982) und IW (1980 a) keine Angaben. Erst dem kaum zugänglichen Projektbericht (IW 1981 c) und der Sekundärliteratur (BUNK 1980; MÖNIKES 1981) ist zu entnehmen, daß jeder Auszubildende Tests und Fragebogen *allein*, ohne Zeitgrenzen, ohne Kontrollmöglichkeit über benutzte Hilfsmittel bearbeitete und daß es ihm überlassen blieb, ob er das Material zurücksandte (Rücklauf 60 Prozent). Das ist bei Leistungstests eine so ungewöhnliche Bedingung, daß man in der Publikation eine Diskussion der Fehlermöglichkeiten erwarten müßte. Es ist ja keineswegs so, daß alle Auszubildenden die damit verbundenen Chancen in gleichem Maße ausnut-

zen oder daß nun alle Testergebnisse (linear?) besser sein müßten als bei standardisierten Durchführungsbedingungen. Jeder erfahrene Testleiter weiß, daß es zum Beispiel Kandidaten gibt, die ihre Arbeit lange vor der Zeitgrenze abgeben wollen und erst durch den Hinweis auf Ausnutzung der Überprüfungsmöglichkeiten zu mehr Punkten gebracht werden. Völlig offen bleibt der Einfluß dieser Maßnahme auf Anpassungstendenzen bei „Arbeitsugenden“ und „Werthaltungen“. Der Aussagewert der vielen bei GÖBEL (1982) publizierten Vergleiche zwischen Auszubildenden mit verschiedener Vorbildung usw. ist aufgrund der fehlenden Standardisierung der Untersuchungsbedingungen völlig ungewiß, die Interpretationen sind als Vermutungen einzustufen.

Wie in allen bisher erwähnten Untersuchungen ist auch hier die Diskussion der *inhaltlichen Gültigkeit* unterblieben oder völlig unbefriedigend. Da reicht es nicht, wenn BUNK (1983) die Elementarkenntnisse in Rechnen und Rechtschreibung als grundlegende Voraussetzung für eine „fundamentierte Allgemeinbildung“ bezeichnet und die als „Ignoranten“ abtut, die auf elektronische Rechner und Textverarbeitungsautomaten verweisen (S. 725). Man kann darüber diskutieren, welche Elementarkenntnisse noch zu den „grundlegenden Voraussetzungen“ zählen (Verständnis bei stillem Lesen z. B.), welche Kenntnisse überhaupt für die Berufsausbildung wichtig sind, aber man kann nicht darüber diskutieren, daß von Rechen- und Rechtschreibaufgaben her keine Aussagen über den Wert von Gesamtschulen oder die Einführung des 10. Schuljahres möglich sind. Genau das geschieht bei den Interpreten dieser Untersuchungen aber immer wieder.

Die Ergebnisse der IW-Studie werden in Prozenttabellen und Grafiken mitgeteilt. Für *Rechnen* gibt es auch eine klare Gesamtaussage: „Im Durchschnitt wurden 24,2 Aufgaben richtig gelöst. Das entspricht einem Prozentsatz von 71,2 Prozent (IW 1981 a, S. 2). Wenn es sich tatsächlich um einen traditionellen Test handeln würde, bei dem ein mittlerer Schwierigkeitsgrad von 50 Prozent angezielt wird, dann könnte das ein sehr gutes Ergebnis sein. Aber nicht standardisierte Untersuchungsbedingungen verbieten auch eine derartige Aussage. Bei der *Rechtschreibung* kann GÖBEL (1982, S. 66f.) entnommen werden, daß 68 Prozent die Grundkenntnisse (16 Aufgaben) besaßen, 45 Prozent die vier Aufgaben der Groß- und Kleinschreibung lösten und 79 Prozent die vier Fremdwortaufgaben. Ein Vergleich mit den von den Testautoren ermittelten Normwerten wird nicht vorgenommen. Stellt man ihn an (wie HAENISCH 1982), so erkennt man, daß die Ergebnisse zum Teil 2 Prozent unter, zum Teil 5 Prozent über der mittleren Norm liegen. Also auch ein gutes Ergebnis, wenn man bei diesen Durchführungsbedingungen überhaupt etwas sagen könnte. Aber die Ausbilder fordern zum Beispiel 83,6 Prozent bis 100 Prozent Sicherheit in der Groß- und Kleinschreibung (!), und so ergibt sich immer wieder die Möglichkeit zu Klagen. Nach den methodischen Standards dieser Untersuchung ist eine weitere Diskussion ihrer Ergebnisse überflüssig.

Wer dieser Untersuchung die Einhaltung wissenschaftlicher Standards bescheinigt, scheint diese Standards nicht zu kennen. BUNK (1982, S. 723), Mitglied des wissenschaftlichen Beirates für diese Untersuchung, geht aber noch weiter: „Der Minister hat die Arbeit des Instituts ausdrücklich anerkannt und die Ergebnisse

akzeptiert. Was nun die Ergebnisse der Hauptschüler betrifft, so stimmen sie ... mit den hier vorgestellten ähnlich gelagerten Kammer-Untersuchungen weitgehend überein. Damit haben auch diese Untersuchungen von offizieller Seite eine indirekte Bestätigung erfahren.“ Diese Methode der Validierung wird sich in den empirischen Sozialwissenschaften sicher nicht durchsetzen.

### 3.2. Wiederholte Querschnitte unter vergleichbaren Bedingungen

Die Beurteilung von Leistungsveränderungen, die in fast allen diesen Untersuchungen bereits erfolgte, setzt wiederholte Querschnittuntersuchungen mit vergleichbaren Stichproben (Kohorten) und Instrumenten voraus. Wissenschaftlich geplante Untersuchungen dieser Art liegen zu diesem Thema in Deutschland nicht vor. Es gibt nur von einigen Großbetrieben fragmentarische Angaben zu Untersuchungen an ihren Bewerbern (anfallende Stichproben) in verschiedenen Jahren. Außerdem sind in Hessen – wie erwähnt – zwei Jahrgänge mit Aufgaben früherer Jahre untersucht worden. Für die Vergleiche in Hessen ergeben sich folgende Daten zur Gesamtstichprobe (o. Autor, Kurhess. Wirtschaft 1981; FREYTAG 1983):

Bei diesen Aufgaben und diesen Stichproben ist keine Veränderung aufgetreten, die praktisch bedeutsam wäre. BUNK (1983) übergeht die Mittelwerte von 1978/1982 und errechnet stattdessen aus den Tabellen Prozentanteile für nicht ausreichende Leistungen bei Hauptschülern, Realschülern und Gymnasiasten und kommt so zu einem weiteren „Beweis“ für Leistungsabfall. Da der relative Anteil der Hauptschüler von 1978 zu 1982 stark abgefallen ist (ALTHOFF/WERNER 1983) und die Hauptschule stärker negativ selektiert ist, kann man zu solchen Zahlen kommen. Allerdings mußte BUNK die Gesamtschüler mit Haupt- und Realabschlüssen weglassen, da sonst der Trend durchbrochen wäre.

BUNK (1983) führt außerdem noch die Untersuchungen der Estel Hüttenwerke Dortmund AG und der Hoechst AG an. Für Estel (Betriebliche Ausbildungspraxis, ohne Autor 1982) heißt es, daß die seit 1969 durchgeführten Kenntnisprüfungen und Begabungstests in ihren Ergebnissen „vergleichbar sind, weil die Aufgabenstellungen u. a. immer auf das gleiche Anspruchsniveau abgestellt sind“ (S. 930). Für Rechnen wurden 18 Aufgaben zum „üblichen Repertoire“ gegeben, deren Schwierigkeitsgrad so gewählt wurde, daß sich bei den Bewerbern „vor rund 10 Jahren eine der Normalverteilungskurve angenäherte Streuung der Benotungen ergeben mußte“ (S. 930, wörtlich auch bei BUNK 1983, S. 724). Sieht man sich die Ergebnisdarstellung an, so erkennt man, daß 1969 die Noten 5 und 6 einen Prozentanteil von 40 Prozent hatten, die Note 4 auf 25 Prozent kam und die

Tab. 3: Wiederholte Querschnitte in hessischen IHK-Untersuchungen

	Rechnen	Rechtschreibung
Mittlere Fehlerzahl 1977	12,5	16,2
Mittlere Fehlerzahl 1980	12,3	17,0
Mittlere Fehlerzahl 1978	14,2	14,0
Mittlere Fehlerzahl 1982	14,2	14,0



Noten 1, 2 und 3 zusammen auf 35 Prozent. Es gibt keine der Normalverteilung angepaßte Zensurenverteilung, bei der 5 und 6 auf 40 Prozent kommen. Der Anteil der Noten 5 und 6 steigt dann 1980 auf 80 Prozent, während 1, 2 und 3 auf unter 10 Prozent abfallen. Dieses Faktum und das Fehlen jeder Angabe zu Gütekriterien und Äquivalenz der Parallelförmigen (?) sollte eigentlich verhindern, daß solche Ergebnisse in wissenschaftliche Abhandlungen als Beweis übernommen werden. Das seit 1973 durchgeführte PSB von HORN zeigt einen (nicht kontinuierlichen) Anstieg von 98 auf 102 Punkte, obwohl in diesem Begabungstest Rechtschreibkenntnisse und Kopfrechnen eine nicht unerhebliche Rolle spielen.

Für Hoechst konnte sich BUNK (1983) auf eine Mitteilung von AMTHAUER beziehen. Dort wurde seit mehr als 25 Jahren ein Rechentest angewandt, „dessen unveränderte Bestandteile die vier Grundrechnungsarten, einfaches Bruchrechnen und einfaches Prozentrechnen sind“ (BUNK 1983, S. 723). Wir erfahren nichts über die Gütekriterien des Tests und wissen nicht, ob seine Aufgaben unverändert oder äquivalent sind oder nur sein Aufbau. Darum ist der Wert der Aussage, daß die Rechenleistungen der Sechzehnjährigen von 1962/65 bis 1976/79 um 13 Prozent nachgelassen hätten, auch nicht zu beurteilen.

Die Wirtschaftsvereinigung Eisen- und Stahlindustrie (1982) hat noch weitere Firmenergebnisse vorgelegt, die sie so zusammenfaßt: „Die Ergebnisse der Rechentests von Ausbildungsbewerbern mit Hauptschulabschluß beweisen jährlich geringer werdende Kenntnisse. Die Intelligenztests beweisen dagegen ein nahezu gleichbleibendes Begabungsniveau“ (S. 2). Zum „Beweis“ werden aber wieder nur Kurvenverläufe ohne Angaben zu den Tests und zu den Stichproben angeführt, die die Zusammenfassung nicht immer stützen.

Uns liegt noch eine Ergebnistabelle der BASF vor, die aus jährlich etwa 8000 Bewerbern nach Zufall die Testergebnisse von je 150 Kandidaten mit Hauptschulabschluß und Mittlerer Reife zusammenstellte. Angewandt wurden der Teil „Rechtschreibungskorrigieren“ (Maximal 40 Punkte) des Allgemeinen-Büro-Arbeits-Tests (A-B-A-T) von LIENERT und ein selbst entwickelter Rechentest mit 20 Aufgaben (Grundrechnarten und Prozentrechnen), dessen Gütekriterien nicht mitgeteilt wurden. Hier werden die Prozentanteile richtiger Lösungen für die Bewerber mit Hauptschulabschluß angeführt.

Für Rechnen ist bis 1974/75 ein unregelmäßiger Verlauf mit der Höchstleistung im Jahr 1973/74 zu erkennen, dann ein Absinken bis 1980/81, auf das dann eine leichte Erholung zu folgen scheint. Bei Rechtschreibung liegt der Höhepunkt 1975/76 und der Tiefpunkt 1982/83, dem wieder ein leichter Anstieg folgen könnte. Aber ohne genauere Daten sind auch diese Trendaussagen mit großem Risiko behaftet.

*Tab. 4: Prozentanteile richtiger Lösungen in Eingangstests der BASF*

Jahr	Rechnen	Rechtschreiben
1970/71	74,05	—
1971/72	73,05	—
1972/73	72,25	—
1973/74	75,2	—
1974/75	72,5	51,1
1975/76	71,2	53,0

Tab. 4 (Fortsetzung): Prozentanteile richtiger Lösungen in Eingangstests der BASF

Jahr	Rechnen	Rechtschreiben
1976/77	69,8	50,1
1977/78	63,8	48,8
1978/79	61,1	49,3
1979/80	57,9	47,3
1980/81	56,5	45,7
1981/82	58,0	44,6
1982/83	57,0	43,7
1983/84	60,9	45,9

### 3.3. Zusammenfassende Stellungnahme zu den Untersuchungen an Berufsanfängern

- a) Es ist verständlich und berechtigt, daß die Ausbildungsbetriebe ein Interesse daran haben, mit welchen Voraussetzungen die Schulabgänger zu ihnen kommen und ob sich diese Voraussetzungen mit der Zeit verändern.
- b) Diese Voraussetzungen für die Berufsausbildung sind in den vorliegenden Untersuchungen mit Rechtschreibung und Rechnen jedoch höchst fragmentarisch erfaßt worden. Eine curriculare Analyse, was gelehrt wurde und was gefordert werden muß, ist in den Untersuchungen nicht einmal im Ansatz erkennbar.
- c) Die angeführten Untersuchungen entsprechen in keinem Falle den international anerkannten Standards empirischer Sozialforschung. Die Defizite bei der Stichprobenbildung, den Untersuchungsinstrumenten, der Datenverarbeitung und Berichterstattung sind eklatant. Welche Wirkungen zum Beispiel der zwischen 1973 bis 1982 von 3,4 auf 9,6 Prozent gestiegene Ausländeranteil an den Schulpflichtigen auf die Rechtschreibergebnisse hat, wird nicht angesprochen, geschweige denn ermittelt.
- d) In krassem Gegensatz dazu steht der Interpretationsdrang der Untersucher. Es gibt kaum einen Aspekt der Schulreform, zu dem nicht dezidiert Stellung genommen wird. Diese Stellungnahmen sind mit beachtlichem Aufwand der Presse unterbreitet und dort meist völlig unkritisch übernommen worden (vgl. DIHT 1967).
- e) Nach wissenschaftlichen Kriterien liegt kein generalisierbarer Beweis für eine Leistungsveränderung in positiver oder negativer Richtung vor. Wir sind auch nicht berechtigt, wie bei subjektiven Erfahrungsberichten, anzunehmen, daß sich Milde- und Strengeeffekte in etwa ausgleichen werden und daß der Fülle der Klagen dann doch Beweiskraft zukommt. Dem widerspricht die weitergreifende historische Analyse, die zur Annahme eines gerichteten Fehlertrends, hier Verklärungseffekt genannt, führte. Dieser Trend konnte bei INGENKAMP 1967b an einer kleinen Stichprobe demonstriert werden.

#### 4. Untersuchungen in Schulen

Mit Ausnahme des später zu diskutierenden Diktatvergleiches aus Rheinland-Pfalz haben die Schulbehörden der Bundesrepublik noch keine Untersuchung zur Leistungsveränderung veranlaßt. Gegen die Vorwürfe „der“ Wirtschaft wehrten sich Pädagogen und Bildungspolitiker mit anderen Interpretationen der vagen Befunde (z. B. MÖNIKES 1981; WITTMANN 1981) oder mit kleineren Nachuntersuchungen (z. B. GLATZ 1967). Diese schwache Reaktion mag damit zusammenhängen, daß die Untersuchungen nicht sehr ernst genommen wurden (dann hätte man sie aber kritisch aufarbeiten können) und/oder daß die Pädagogen zu einem erheblichen Teil selbst an einen Leistungsrückgang glauben. Haben wir denn empirische Befunde zur Überprüfung dieser Aussagen?

An anderer Stelle (INGENKAMP 1967b) wurde bereits ausführlich begründet, warum frühere Zensuren- und Eignungstest-Vergleiche von W. HARTNACKE, K. V. MÜLLER und A. HUTH hier nicht heranzuziehen sind. Dort wurde auch über die einzige Wiederholung einer Querschnittuntersuchung in deutschen Schulen berichtet, die die damals gültigen Kriterien pädagogischer Tatsachenforschung erfüllte.

Heute noch interessant an dieser Untersuchung ist, daß sie einen Vergleich der Lehrmeinung über Leistungsveränderungen mit den Testergebnissen ermöglichte. Die Untersuchungen fanden 1949 und 1962 in allen 6. Klassen eines Berliner Bezirks ( $n = 1200-1400$ ) statt. Die Nachkriegs-Schulverhältnisse waren 1949 ungleich schlechter. Jene 20 Lehrer, die in beiden Jahren 6. Klassen unterrichteten, hatten 1949 nur 6. Volksschulklassen (eine Negativauslese), 1962 aber die oberste Klasse der nunmehr sechsjährigen Berliner Grundschule unterrichtet. Alles sprach eigentlich dafür, eine Verbesserung der Schulleistungen anzunehmen. Die Lehrer glaubten dennoch in den Bereichen Rechtschreibung, Bruchrechnen und Verständnis bei stillem Lesen nur zu 5–20 Prozent an eine Leistungsverbesserung, sahen dagegen mit 35–50 Prozent eine Verschlechterung. Tatsächlich sind die Mittelwerte in allen Tests sehr signifikant angestiegen. Diese Leistungsverbesserung war – wie gesagt – durch die Überwindung der Nachkriegsprobleme zu erwarten. Sie kann nicht auf spätere Jahre übertragen werden. Wichtig ist jedoch der Nachweis, daß die überwiegend negative Veränderungsschätzung der beteiligten Lehrer kein Indiz für die tatsächliche Leistungsveränderung sein muß, sondern im Widerspruch zu den Fakten stehen kann. Dieser Nachweis stützt die These vom Verklärungseffekt.

Was sonst aus der Schule zu berichten ist, trägt teilweise den Charakter der Kuriosität (vgl. INGENKAMP 1984). Im Anschluß an eine Mainzer Tagung (*Symposium* 1984) gab das Kultusministerium Rheinland-Pfalz 1984 „Empfehlungen zur Rechtschreiberziehung“ (KM 1984) heraus. Im Vorwort erklärte der Kultusminister, daß zur Veränderung der Rechtschreibleistungen sehr unterschiedliche Auffassungen vertreten würden. „Ich habe daher eine Überprüfung an ausgewählten Grund- und Hauptschulen, Realschulen und Gymnasien veranlaßt.“ Zu den Vergleichsdiktaten wird mitgeteilt, daß sie 1984 in etwa 10 Prozent der Klassen durchgeführt wurden. Als Ergebnis wird festgestellt: „In der Grundschule wurde das gleiche Diktat geschrieben wie bei der landesweiten Erhebung über den Leistungs- und Begabungsstand der Volksschüler. Der Vergleich

... von 1959 und 1984 zeigt, daß die Rechtschreibleistungen in den Grundschulen heute nicht schlechter sind als 1959 im 4. Schuljahr der Volksschule.“ Für 1984 wird noch mitgeteilt, daß 4390 Schüler aus 213 Klassen das Diktat mit 107 Wörtern schrieben und durchschnittlich 4,7 Fehler machten.

Alle Daten zu 1959 fehlen ebenso wie Angaben zur Stichprobenbildung, zur Erprobung des Diktats, zur Signifikanz von Unterschieden. Die Standards empirischer Forschung werden noch krasser verletzt als bei den Untersuchungen der IHKn. Die Angaben zu den anderen Schultypen, die nur 1984 untersucht wurden, sind nicht gehaltvoller. Die zusammenfassende Feststellung: „Die Überprüfung widerlegt eine pauschale Kritik an den Rechtschreibleistungen unserer Schüler. Die Leistungen sind insgesamt zufriedenstellend“ ist trotz ihres Pressechos eine unbewiesene Meinungsäußerung.

Zusammenfassend ist festzustellen, daß das staatliche deutsche Schulwesen keine vom Lehrerurteil unabhängigen Daten über den Lernerfolg seiner Absolventen vorlegen kann. Die Kultusministerien verweigern darüber hinausgehend auch Forschern den Schulzugang, wenn der Verdacht besteht, es könnten objektive Daten über den Erfolg schulischer Maßnahmen erhoben werden (DGfE 1980; HITPASS 1985; INGENKAMP 1980). Die Untersuchungen, die von den „Abnehmern“ des Schulsystems vorgelegt werden, zeigen, daß ein Bedarf für überprüfbare Informationen besteht. Auf Untersuchungsergebnisse aus diesem Kreis haben die Schulverwaltungen bisher nur mit dem Versuch der Uminterpretation oder völlig dilettantischen Eigenerhebungen geantwortet. Diese Reaktion unterscheidet sich deutlich von der Praxis anderer westlicher Industrienationen, die SCHREIBER in diesem Heft beschreibt.

### *5. Untersuchungen an Studienanfängern*

Über die Anforderungen an eine „Hochschulreife“ kann hier nicht berichtet werden (vgl. dazu u. a. HELDMANN 1984). Auch die Umfragen unter Hochschullehrern sollen nur am Rande erwähnt werden, da sie nur Aussagen über Leistungserwartungen erlauben. Die umfangreichste Studie mit den Antworten von 1279 Mitgliedern des Hochschulverbandes (Rücklaufquote 11,35 Prozent) hat jüngst HELDMANN (1984) vorgelegt. Die Hochschullehrer haben Ausbildungsinteresse, Denkvermögen sowie Lern- und Leistungsbereitschaft „relativ positiv“ (S. 382) eingeschätzt, während „die Arbeitsqualität, das Ausdrucksvermögen, die Technik der schriftlichen Darstellung, die Präsenz des Wissens, die Selbständigkeit und die Motivation, die Ausdauer und die Belastbarkeit und schließlich das Differenzierungsvermögen unzureichend entwickelt worden sind“ (S. 383). HELDMANN fordert auf, künftig „objektive Daten zu erheben, die dann als Kontrolldaten ... dienen können“ (S. 376). HITPASS (1985) hat 6957 Studierende (35,2 Prozent der Angefragten) nach den Merkmalen der Studierfähigkeit und der Ausprägung dieser Merkmale bei ihnen selbst befragt. Die erfreuliche subjektive Kritik in der Selbsteinschätzung wird leider bei ihm zu „einem doch bemerkenswerten Defizit an allgemeiner Studierfähigkeit“ (S. 45) uminterpretiert.

Diese beiden jüngeren, größeren Umfragen kommen zu eher moderaten, ausgleichenden Bewertungen und sind in ihren Schuldzuweisungen vorsichtig und abwägend. Das unterscheidet sie deutlich von den vielen Einzeläußerungen, für die nur GAUS-FALTINGS u. a. (1980a und b, S. 1) angeführt werden sollen: „Aufgrund eigener Lehrerfahrungen und aus Gesprächen mit Fachkollegen anderer Hochschulen läßt sich jedoch deutlich feststellen, daß die Vorkenntnisse der Studienanfänger in den letzten Jahren immer schlechter und – bedingt durch die Einführung des Kurssystems in der reformierten Oberstufe – auch heterogener geworden sind.“

Wortführer der Schuldzuweisung an die neugestaltete gymnasiale Oberstufe war HITPASS (1980, 1981). HAENISCH 1980 hat nachgewiesen, daß die HITPASS-Studie nicht aussagekräftig ist. HITPASS kommt in einer neueren Untersuchung selbst zu dem Schluß, daß die Verdoppelung der Abiturientenanteile von 5 auf 10 Prozent „trotz der einschneidenden Reform der gymnasialen Oberstufe keine gravierenden Veränderungen in der Entwicklung der allgemeinen Hochschulreife erkennen“ läßt (1985, S. 7).

Der Untersuchungsstand zu den Leistungen der Studienanfänger ist in *Tab. 5* und *6* zusammengefaßt. Auch hier haben wir keine einzige Studie, die generalisierende Aussagen über Leistungsveränderungen auch nur in einem Studienfach zulassen würde. Die Untersuchungen in Mathematik für Nicht-Mathematiker sind am zahlreichsten, geben jedoch keinerlei Basis für die Behauptungen der mathematisch-naturwissenschaftlichen Vereinigungen, die unter der Kapitelüberschrift „Verfall der Kenntnisse“ geschrieben hatten: „Langjährige Leistungsmessungen an vielen Hochschulen belegen diese Tatsache“ (REISS/STEINER 1984, S. 57). Hierzu sind der gründliche Beitrag von STEINER (1984) und die Dokumentation der Untersuchungen bei KURZ (1982, 1983) heranzuziehen. Bei den beiden wiederholten Querschnitten mit den größten Stichproben (SCHOENES 1983; LEHMBERG u. a. 1982) ist der Leistungsabfall nicht zu erkennen oder wird in den letzten Erhebungen wieder aufgefangen.

Tab. 5: Untersuchungen an Studienanfängern – Einmalige Querschnitte –

Autor/ Jahr	Region	Stichprobe	n	Angabe zur Repräsen- tanz	Untersuchungs- bereiche	Verfahren	Angabe zu Gütekrite- rien	Ergebnisse	Prüfstu- fistik	Zeitraum
Kurz 1981a, b	24 FHS, da- von 16 aus Baden- Württem- berg	Studien- anfänger in (techn. Richtung)	2048	K. A.	Mathematik, Physik	62 Testaufgaben Mathematik, 55 Physik	+	Im Mittel 39,7% richtig gelöste Mathematikaufgaben, 50,1% Physikaufgaben	K. A.	WS 1979/80
Büning u. Recke 1984	Berlin FU	Studien- anfänger in Wirtschafts- wissenschaft	127	K. A.	Mathematik	26 Mathematik- aufgaben	K. A.	73,2% nicht ausreichende Kenntnisse	K. A.	1980
Bendlin 1985	Uni Hanno- ver	Studienan- fänger der Ingenieur- wissensch.	526	K. A.	Mathematik	11 Rechenaufgaben	K. A.	3,8% Aufg. v. Abiturienten mit Leistungsfach Math. ge- löst, 1,6% Aufg. von Abitu- rienten ohne Leistungsfach	K. A.	WS 1984/85
Krause u. Reiners- Logothet- idou 1979/81	Bundesrepu- blik	Studien- anfänger Physik aus 39 Fach- bereichen	2897	nicht reprä- sentativ	Mathematik, Physik	47 Mathematik- aufgaben, 47 Physik- aufgaben	K. A.	16-30% weisen einen unzu- länglichen Kenntnis- und Fähigkeitsstand auf	K. A.	1978
Leppig 1978	NRW 12 Unis	Studienbe- werber im math. Vorse- mester	769	K. A.	Mathematische Beweisverfahren	3 Aufgaben	K. A.	Leistungen fallen keineswegs so aus, wie es von Studien- bewerbern für Math. erhofft werden mag (S. 103)	K. A.	1976

K. A. = Keine Angaben

Tab. 6: Untersuchungen an Studienanfängern – Wiederholte Querschnitte –

Autor/ Jahr	Region	Stichprobe	n	Angabe zur Repräsen- tanz	Untersuchungs- bereiche	Verfahren	Angabe zu Gütekrite- rien	Ergebnisse	Prüfstu- tistik	Zeitraum
Nägerl u. Harten 1980	Uni Göttin- gen	Erstsemester Medizin vor Physikal. Praktikum	je Jahr mind. 210	K. A.	Mathematik	Mathematiktest für Abiturienten und Studenten von LIE- NERT u. HOFEN, Hogrefe 1972	Rel. = .90 (Parallel- test)	1972–76 Ergebnisse 8 Punkte über Normstichprobe. Ab 1977 bis 1979 ss. Abfall bis rund 3 Punkte unter Norm	Abfall ss	1972 bis 1979
Schoenes 1983 (s. a. Kurz 1983)	FH Darm- stadt	Studienan- fänger im FB Mathe. u. Naturwissen- schaften	3640	K. A.	Mathematik, Physik	20 Mathematikaufg. 20 Physikaufgaben	K. A.	„Keine deutliche Tendenz in Richtung einer Änderung der Eingangskennnisse... weder eine Verbesserung noch eine Verschlechterung.“	K. A.	1977/78 bis 1983/84
Lehmberg u. a. 1982	TH Darm- stadt	Studienan- fänger Bau- und Vernet- zungsinge- nieure	1120	K. A.	Physik	21 Physikaufgaben	K. A.	„... der mittlere Erfolg... zeigt eine ... leicht sinkende Tendenz (jedoch Zunahme von 1981/82 auf 1982/83 nach KURZ 1983).“	K. A.	1974 bis 1980
Schwarze 1985	TU Braun- schweig	Studienan- fänger Wirt- schaftswis- senschaft	K. A.	K. A.	Mathematik	Eigener Eingangstest mit 27 Aufgaben A.-A.-Form	K. A.	1976 durchschn. 64% richtig, 1980 54% und 1983 47%	K. A.	1976, 1980, 1983

Tab. 6: - Fortsetzung

Aut./ Jahr	Region	Stichprobe	n	Angabe zur Reprä.	Untersuchungs- bereiche	Verfahren	Angabe zu Gütekrite- rien	Ergebnisse	Prüfstä- tistik	Zeitraum
Klein- Braley 1981	Uni Duis- burg	Studienan- fänger in Englisch	324 von 1977- 1981	K. A.	Englisch	Placementtest Delta	Rel. = 93 Faktor., Konstrukt- und Über- einst.-Vali- dität	Mittl. Punktwerte sinken in fast allen Untertests	K. A.	Jedes WS von 1977/78- 1980/81
Jahn 1982	Unis Augsburg, Bonn, Köln, Wuppertal	Studienan- fänger	1) 592 bzw. 372 2) 530	K. A.	Englisch	Placementtest H 1 140 Antwort-Aus- wahl-Aufgaben	Rel. = 95 Progn. Vali- dität wahr- scheinlich	1981 4,2 bzw. 2,4 Punkte besser	K. A.	1975 und 1981
Klein- Braley 1983/84	Uni Duis- burg	Studienan- fänger in Englisch	346 von 1982-85	K. A.	Englisch	Placementtest Delta	S. o. 1981, KLEIN- BRALEY/ LÜCK 1980, SCHWIB- BE/ SCHWIBBE 1986	Gleichbleibende Gesamtwerte	K. A.	WS 1982/83 bis SS 1985
Bonheim 1984	Bundesrepu- blik	Bewerber für DAAD	K. A.	K. A.	Englisch	Auslesetest DEST	Rel. = 95 Faktorielle Validität	Uneinh. Abfall bis 1980, dann konstanter Anstieg	K. A.	1976 bis 1984

K. A. = keine Angabe



BERCHEM (1984, S. 5) sprach demgegenüber von „alarmierenden Ergebnissen“ von Studienanfängertests. Ohne Literaturhinweise bezog er sich auf Göttingen, Hannover, Braunschweig und Köln. In Göttingen haben NÄGERL und HARTEN (1980) von 1972 bis 1979 die Erstsemester in Medizin vor dem Physikalischen Praktikum mit dem M-T-A-S untersucht. Bis 1976 lagen die Ergebnisse bei etwa 65 Punkten im Mittel, ab 1977 fielen sie bis auf etwa 54 Punkte. Dies ist sicher nicht der von BERCHEM behauptete Rückgang „um ein Drittel“, und vor allem sagt das Ergebnis nicht viel, da die Autoren die möglicherweise über die Jahre nachlassende curriculare Gültigkeit des Tests und mögliche Veränderungen in der Stichprobenzusammensetzung nicht überprüft haben. Aus Hannover gibt es eine Kurzmitteilung von BENDLIN 1985 über die Ergebnisse unerprobter Rechenaufgaben, die TIETZ seinen Studenten vorlegte. Sie hat eher anekdotischen Wert. Aus Braunschweig liegt ein kurzer Bericht von SCHWARZE (1985) vor, der angehenden Wirtschaftswissenschaftlern 27 Aufgaben zur Schulmathematik vorlegte, aber nichts über die Gütekriterien des „Tests“, Stichprobengröße und -veränderung aussagt und daher keine Beweiskraft beanspruchen kann. Zu der Kölner Untersuchung gibt es nur eine Pressemitteilung der Fakultät und eine Kurzbeschreibung bei KURZ (1983, S. 114/115), die erkennen läßt, daß diese informelle Erhebung nicht für generalisierende Aussagen geplant war und solche auch nicht zuläßt. Aussagekräftiger wären die umfangreichen und sorgfältigen Untersuchungen an Fachhochschulen, besonders in Baden-Württemberg, gewesen (KURZ 1981; KURZ u. a. 1981), zu denen KURZ (1981, S. 19) feststellt: „Wie die dargestellten Testergebnisse zeigen, ist das eigentliche Problem *nicht*, daß die Vorkenntnisse im Durchschnitt zu niedrig liegen, sondern offensichtlich sind die Vorkenntnisse der Studienanfänger sehr heterogen.“

BERCHEM (1984, S. 5) bezieht sich außerdem auf einen Test der Deutschen Physikalischen Gesellschaft, nach dem die Erstsemester im Studienfach Physik nur die Hälfte der gestellten Mathematikaufgaben und nur 42 Prozent der Physikaufgaben lösten, obwohl „sie nur elementare Kenntnisse voraussetzen“ sollten (KRAUSE/REINERS-LOGOTHETIDOU 1979, 1981a). Der Test bestand aus je 47 Aufgaben für Mathematik und Physik und hatte eine relativ starke Speed-Komponente (20 Prozent der Teilnehmer wurden nicht fertig). Jegliche Angabe über die curriculare Validität des Tests oder seine Gütekriterien fehlt. Der Test wurde 1978 in 39 von 45 Fachbereichen Physik bundesweit in einer Querschnittuntersuchung eingesetzt ( $n = 2718$ ).

Die Ergebnisse werden in graphischer und beschreibender Darstellung sowie mit prozentualen Anteilen für typische Lösungen mitgeteilt. Da keinerlei Vergleichsdaten aus vergangenen Jahren vorliegen, bleibt nur der Eindruck einer „Diskrepanz zwischen dem, was allgemein – insbesondere auch von vielen Hochschullehrern – von den Physikanfängern erwartet wird und dem, was viele Physikanfänger tatsächlich zu leisten vermögen“ (1981a, S. 298). Diese Diskrepanz wird im umfangreichen Abschlußbericht detailliert aufgearbeitet, gleichzeitig immunisieren die Autoren ihr Vorgehen gegen Kritik mit dem Hinweis, daß „die Zielsetzung der herkömmlichen Testtheorie nicht mit der Zielsetzung unseres Tests“ übereinstimmt. „Eventuelle Vorhaltungen ‚methodischer Art‘ gegenüber unserer Arbeit ... könnten nicht einfach durch Verweis auf die ‚üblichen‘ Methoden gerechtfertigt werden“ (KRAUSE/REINERS-LOGOTHETIDOU 1981b, S. 77). Da die Autoren auch zeitlich nicht in der Lage waren, sich in neuere Testtheorien und -methoden einzuarbeiten, bleibt – außer Mißverständnissen – zur Begründung ihres eigenen Weges vor allem das Unbehagen des theoretischen Physikers an der „Datenreduktion“ durch Testtheoretiker (S. 45f.).

Die *Tabelle 6* enthält auch Angaben zu einigen wiederholten Querschnitten in Englisch. Obwohl BERCHEM (1984, S. 4) anführt, daß „fehlende oder unzureichende Fremdsprachenkenntnisse bemängelt“ werden, geht er auf keine dieser Studien ein. Sie passen nicht in das Negativbild, zeigen teils abfallende, teils gleichbleibende, teils ansteigende Tendenz, wie man es bei kleineren anfallenden Stichproben erwartet. Wenn der Präsident der WRK weitreichende Reformvorschläge zur Umgestaltung des Gymnasiums macht und als entscheidende Begründung einige „Untersuchungen“ herausgreift, die zum Teil nur in Form von Pressemitteilungen zugänglich wurden, und sie als Indiz für tatsächliche Veränderungen nimmt, dann deutet das an, wie wenig elementare Grundregeln sozialwissenschaftlicher Methodik in Bildungsplanung und -politik Beachtung finden.

Zusammenfassend ergibt sich auch für die Gruppe der Studienanfänger keine Untersuchung, die eine generalisierende Aussage über Leistungsveränderungen zulassen würde. Die vorliegenden Studien stammen fast nur aus den Bereichen Mathematik und Englisch. Im Gegensatz zur Erwartung zeigen Mathematikuntersuchungen häufig die gravierendsten Verstöße gegen anerkannte Regeln empirischer Sozialforschung. Die Untersuchungsinstrumente sind meist unüberprüfte subjektiv ausgewählte Aufgaben aus der Sekundarstufe I. Die Interpretationswillkür ist bestürzend, wird aber in der eigenen Fachliteratur erkannt und korrigiert (z. B. SCHMIDT 1982; STEINER 1984). Die wenigen Untersuchungen in Englisch mit zum Teil beachtenswerten Tests lassen keinen einseitig absinkenden Trend erkennen, gestatten aber auch keine Generalisierung. Die beachtlichste Arbeit in diesem Bereich überhaupt, die vom BMBW geförderte Physik-

studie, hat der Aussagekraft ihres Querschnitts und den Möglichkeiten der Wiederholung selbst enge Grenzen gesetzt, weil die internationalen Erfahrungen mit neueren Testtheorien und Testprogrammen nicht aufgearbeitet werden konnten.

## 6. Resümee

Die Durchsicht der vorliegenden Untersuchungen hat gezeigt, daß wir zur Leistungsveränderung unserer Berufs- und Studienanfänger keine hinreichend tragfähige Aussage machen können. Wir haben Hinweise auf absinkende Leistungen, auf gleichbleibende Lernerfolge, ja sogar auf Lernzuwachs, wenn man die vagen Daten nur ins rechte Licht rückt. Für politische Schaukämpfe ist das Material „ergiebig“ genug und auch weidlich dafür genutzt worden. Aber für eine wissenschaftlich verantwortbare Trendanalyse sind fast alle Ergebnisse viel zu fragwürdig.

Man könnte der nachwachsenden Jugend keinen Vorwurf machen, wenn sie die sogenannten „Kulturtechniken“ weniger beherrscht, denn wir haben ja die Lernumwelt mit Bildmedien, Reizüberflutung, Oberflächlichkeit usw. so gestaltet, daß der Wert dieser „Kulturtechniken“ immer weniger einsichtig wird. Aber wir müssen uns hüten, aus unseren impliziten kulturellen Verfallstheorien und aus Verklärungseffekten vorschnell ein Erklärungsmodell zu konstruieren und ihm die Fakten anzupassen. Gerade der geschichtliche Rückblick sollte uns die Notwendigkeit von Untersuchungen beweisen, die gegen subjektive Verzerrung möglichst abgesichert sind.

Und zu den vorgestellten Untersuchungsansätzen ist nun eine gesicherte Aussage möglich: Fast alle überprüften Untersuchungen verletzen die Standards empirischer sozialwissenschaftlicher Forschung in so schwerem Umfang, daß sie nahezu wertlos sind. Die methodische Diskussion des letzten halben Jahrhunderts findet in fast allen Untersuchungen nicht den geringsten Niederschlag. Wollte man sich der Sprache der Berichterstatter bedienen, so könnte man nur Vokabeln wie ungenügend, katastrophal, dilettantisch, ja sogar demagogisch benutzen. Erstaunlich ist, daß Berufsgruppen, bei denen Sachlichkeit und Faktengenauigkeit assoziiert werden, in diesen Untersuchungen besonders durch Verzicht auf Absicherung gegen subjektive Verzerrung und durch naive Überinterpretation auffallen.

Die Diskussion um die Leistungsveränderungen unserer Berufs- und Studienanfänger hat eminente praktische Relevanz, wie das ständige öffentliche Interesse beweist. Aber die empirisch orientierten Sozialwissenschaften haben sich um diese Frage bisher nicht gekümmert. Kaum ein namhafter Vertreter der Pädagogik, Psychologie oder Soziologie hat sich an diesen Untersuchungen beteiligt, in den akademischen Publikationen existieren sie praktisch nicht. Die Situation ist sehr ähnlich wie in der Klassenfrequenz-Forschung (INGENKAMP u. a. 1985): Hohe praktische Relevanz korrespondiert mit geringem akademischen Stellenwert. Wie der Beitrag von W.H.SCHREIBER in diesem Heft beweist, ist das in anderen Ländern anders und hat nicht nur zu einer Fülle tragfähiger

Daten geführt, sondern auch zu einem differenzierten methodischen Instrumentarium, das bei uns überhaupt nicht genutzt wurde.

Aus Raumgründen verbietet sich hier eine Aufnahme der Diskussion, ob, in welchem Maße und in welcher Form die Gesellschaft von der Schule Rechenschaft über ihre Lehrerfolge erwarten darf. Daß ständig versucht wird, die Unterrichtseffektivität der Schule zu überprüfen, beweisen die hier angeführten Untersuchungen. Daß Öffentlichkeit und Bildungspolitik daran interessiert sind, beweist die öffentliche Resonanz. Vor diesem Hintergrund wirken die bei uns üblichen, hier nur sehr knapp skizzierten Verhaltensmuster nicht sehr überzeugend: 1. Wir diskutieren die Frage einer Rechenschaftspflicht der Schule gegenüber der Gesellschaft (*accountability*) nicht aus und verweisen auf Autonomie und Primat der Erziehung. 2. Wir kommen nicht zu grundsätzlichen curricularen Übereinkünften und/oder Abgrenzungen zwischen abgebenden und aufnehmenden Institutionen und geraten dadurch in Dauerkonflikte. 3. Wir bestreiten einerseits, daß die wichtigsten Lern- und Erziehungsziele unserer Schule meßbar seien, und überprüfen andererseits inhaltlich und methodisch unreflektiert sogenannte „Kulturtechniken“. Angesichts der Intensität, mit der curriculare und methodische Diskussionen vor allem in angelsächsischen Ländern geführt worden sind, wirkt der bei uns beliebte Diskussionsersatz mit Hinweisen auf den „vermessenen Menschen“ ebenso naiv wie arrogant. Veränderungen im Niveau und/oder den Schwerpunkten des schulischen Lernens sind für eine moderne Industriegesellschaft von außerordentlicher Bedeutung. Nicht nur die materielle, auch die kulturelle Existenz einer Nation hängt davon ab, mit welchem Ernst sie über ihre Erziehungsziele und -erfolge diskutiert. Niemand wird sich der Illusion hingeben, eine Fülle wissenschaftlich gesicherter Daten würde alle Fragen beantworten oder uns gar die Zielentscheidung abnehmen. Aber es ist nicht nur für die Glaubwürdigkeit unserer Diskussion, sondern auch für die Qualität unserer Zukunftsplanungen ein Unterschied, ob wir von unzuverlässigen Datenfragmenten ausgehen oder von einem hinsichtlich seiner Aussagekraft überprüften Satz repräsentativer und zu aktualisierender Befunde. Wir haben die Möglichkeiten, die uns die empirische Sozialwissenschaft dazu heute bietet, in der deutschen Pädagogik und Bildungspolitik noch nicht einmal zur Kenntnis genommen.

### *Anmerkung*

1 Im Zentrum für empirische pädagogische Forschung der Erziehungswissenschaftlichen Hochschule Rheinland-Pfalz, Landau/Pfalz, wird gegenwärtig ein Projekt bearbeitet, in dem Ziele, Methoden und Ergebnisse von überregionalen und/oder wiederholten Lernerfolgsmessungen im internationalen Vergleich analysiert werden. Dieser Beitrag gibt einen Überblick über Methoden und Ergebnisse deutscher Untersuchungen an den Schnittpunkten „Berufs- und Studienanfänger“. Der Beitrag von W.H.SCHREIBER in diesem Heft informiert über Untersuchungsprogramme vor allem im angelsächsischen Raum.

*Literatur*

- ALBERT, W.: Der Leistungsrückgang in den Schulen. In: Schule und Leben 4 (1953), H. 10/11, S. 332–343.
- ALTHOFF, H./WERNER, R.: Die Entwicklung der schulischen Vorbildung der Auszubildenden und ihre Auswirkungen. In: Zeitschrift für Berufs- und Wirtschaftspädagogik 79 (1983), S. 122–129.
- ARNHOLD, C./SENFT, F.: Schulische Voraussetzungen beruflicher Ausbildung. In: Arbeitsschulung 1 (1930), Nr. 3, S. 4–10.
- BENDLIN, R.: Das Fach Mathematik darf nicht auf Sparflamme gekocht werden. In: Uni intern (Universität Hannover) 12 (1985), H. 1, S. 13.
- BERCHEM, T.: Was erwarten die Universitäten vom Gymnasium und seinen Abiturienten? Vortrag auf der Akademie Tutzing. Bonn: WRK 1984 (vervielfältigt).
- BERNHEIM, F.: Der Universitätsunterricht und die Erfordernisse der Gegenwart. Berlin/Greifswald 1898.
- BMBW: Informationen – Bildung und Wissenschaft. (1984), H. 10, S. 171.
- BORTZ, J.: Lehrbuch der empirischen Forschung. Berlin 1984.
- BRAND, P./MAISCH, U.: Erster Berufsanfängertest. In: Wirtschaftliche Nachrichten der IHK zu Aachen o. Jg. (1973), S. 560–564.
- BÜNING, H./RECKE, C.: Eingangstests und Brückenkurs in Mathematik im WS 1983/84 bei Studienanfängern im Fachbereich Wirtschaftswissenschaft der FU Berlin. Diskussionsarbeit Nr. 4/1984. Institut für Quantitative Ökonomie und Statistik.
- BUNK, G.P.: Feststellungen der Rechen- und Rechtschreibkenntnisse des Berufsnachwuchses durch die Wirtschaft. (Materialien zur Arbeits-, Berufs- und Wirtschaftspädagogik, Folge 6.) Gießen 1980.
- BUNK, G.P.: Zur Leistungsfähigkeit der Jugendlichen in Rechnen und Rechtschreiben beim Übergang von der allgemeinbildenden Schule in die Berufsausbildung. In: Pädagogische Rundschau 38 (1983), S. 719–739.
- DGE: Die Behinderung der erziehungswissenschaftlichen Forschung in der Bundesrepublik – eine Dokumentation. Tübingen 1980 (Manuskriptdruck).
- DIHT: Der Leistungsstand der Volksschüler bei Beginn der Ausbildung. (DIHT-Schriftenreihe, Heft 99.) Bonn 1966.
- DIHT: Salzpräzeln, Luftbalons und Ohrgelmusik. Was Volksschüler in die Lehre mitbringen. Der DIHT-Test und sein Echo in der Presse. Beuel 1967.
- FREYTAG, H.-P.: Weiterhin erhebliche Lücken in den Grundlagen. In: Wirtschaft und Berufs-Erziehung 30 (1978), S. 210–214. Identisch mit dem Beitrag des Autors in: Kurhessische Wirtschaft (1978), H. 2, S. 51–53.
- FREYTAG, H.-P.: Ergebnisse des Deutsch- und Rechentests 1978. In: Wirtschaft und Berufs-Erziehung 31 (1979), S. 39–43. Identisch mit dem Beitrag „Wieviel Minuten sind 8,25 Stunden?“ des gleichen Autors in: Kurhessische Wirtschaft (1978), H. 1, S. 10–12.
- FREYTAG, H.-P.: Besser als ihr Ruf? In: Kurhessische Wirtschaft (IHK Kassel) 36 (1981), S. 452.
- FREYTAG, H.-P.: Wichtige Grundkenntnisse weiter mangelhaft. In: Kurhessische Wirtschaft (1982), H. 1, S. 10–12. Weitgehend textidentisch mit Runne (1982).
- FREYTAG, H.-P.: Grundkenntnisse nicht verbessert. In: Kurhessische Wirtschaft (1983), H. 1, S. 14–16 (Abdruck auch in Wirtschaft und Berufs-Erziehung 35, 1983, S. 74–78.)
- FURCK, C.L.: Das pädagogische Problem der Leistung in der Schule. Weinheim 1975.
- GAUSFALTINGS, E./MUNDLOS, B./SCHWARZE, J.: Ergebnisse einer Befragung von Dozenten für „Mathematik für Wirtschaftswissenschaftler“ über Mathematikkenntnisse von Studienanfängern. B80-05-01. T.U. Braunschweig: Lehrstuhl für Statistik und Ökonometrie 1980 (vervielfältigt). (a).

- GAUS-FALTINGS, E./MUNDLOS, B./SCHWARZE, J.: Ergebnisse einer Befragung von Studienanfängern der Wirtschaftswissenschaften in Lehrveranstaltungen „Mathematik für Wirtschaftswissenschaftler“ über Mathematikkenntnisse von Studienanfängern. B. 80-06-01. T.U. Braunschweig: Lehrstuhl für Statistik und Ökonometrie 1980 (vielfältigt). (b)
- GLATZ, G.: Bestürzende Unkenntnis? In: Niedersächsische Lehrerzeitung (1967), H. 3, S. 10–14.
- GÖBEL, U.: Was Ausbilder fordern – was Schüler leisten? Köln 1982.
- GRÜNER, G.: „Leistungen im Deutschunterricht die allerungünstigsten.“ In: Die Berufsbildende Schule 36 (1984), S. 533–537.
- HAENISCH, H.: Schlechtes Abitur? – Abiturienten aus Gesamtschule und reformierter gymnasialer Oberstufe. In: Neue Unterrichtspraxis 13 (1980), S. 256–266.
- HAENISCH, H.: Stellungnahme zum Abschlußbericht des Instituts der Deutschen Wirtschaft. Internes Manuskript. Neuss 1982.
- HEILANDT, A.: Berufsanforderungen und Vorbildung – Deutsch und Rechnen in der Volksschule. In: Technische Erziehung 4 (1929), S. 88–89, S. 99–102.
- HELDMANN, W.: Studierfähigkeit. Göttingen 1984.
- HERBIG, M.: Praxis lehrzielorientierter Tests. Düsseldorf 1976.
- HITPAß, J.: Gesamtschule oder Gymnasium? Ein Leistungsvergleich ihrer Abiturienten. Stuttgart 1980.
- HITPAß, J.: Deutschlands Bildungswesen. Die Folge der Reform. Köln 1981.
- HITPAß, J.: Reformierte Oberstufe – besser als ihr Ruf? St. Augustin 1985.
- HOPF, D.: Mathematikunterricht. Eine empirische Untersuchung zur Didaktik und Unterrichtsmethode in der 7. Klasse des Gymnasiums. Stuttgart 1980.
- IHK Bonn: Eine ernste Warnung! Schriftenreihe der IHK Bonn: Heft 9: Lehrabschluß-Prüfung Herbst 1953. Bad Godesberg o. J.
- IHK Essen u. a.: Leistungsstand der Lehranfänger. In: Wirtschaftliche Nachrichten der IHK für die Stadtkreise Essen, Mülheim (Ruhr) und Oberhausen o. Jg. (1968), S. 127–129.
- IHK Hanau: IHK-Test für Berufsanfänger. In: Wirtschaft und Berufs-Erziehung 32 (1980), S. 124–125.
- INGENKAMP, K.: Die deutschen Schulleistungstests. Weinheim 1962.
- INGENKAMP, K.: Zur Methode einer Untersuchung über den Leistungsstand der Volksschüler. In: Pädagogische Rundschau 21 (1967), S. 685–693. (a)
- INGENKAMP, K.: Schulleistungen damals und heute. Weinheim 1967. (b)
- INGENKAMP, K.: Die Freiheit der pädagogisch-psychologischen Forschung – Verfassungsauftrag und Realität. In: AVENARIUS, H./INGENKAMP, K./OTTO, G. (Hrsg.): Forschung und Lehre sind frei ... Weinheim 1980.
- INGENKAMP, K.: Dorothee Wilms Radio-Diktat. In: betrifft: erziehung 17 (1984), S. 10–11.
- INGENKAMP, K.: Lehrbuch der Pädagogischen Diagnostik. Weinheim 1985.
- INGENKAMP, K./PETILLON, H./WEIß, M., u. a.: Klassengröße: Je kleiner, desto besser? Weinheim 1985.
- INSTITUT DER DEUTSCHEN WIRTSCHAFT (IW): Projekt: Überprüfung von Qualifikationsprofilen des Sekundarbereiches I in bezug auf die Qualifikationsanforderungen der Ausbildungsbetriebe für berufliche Ausbildungsgänge. Kurzfassung der wichtigsten Forschungsergebnisse. Bonn 1981 (Manuskriptdruck). (a)
- INSTITUT DER DEUTSCHEN WIRTSCHAFT (IW): Das Kultur-Defizit. In: Informationsdienst des Instituts der Deutschen Wirtschaft 7 (1981), S. 4–5. (b)
- INSTITUT DER DEUTSCHEN WIRTSCHAFT (IW): Überprüfung von Qualifikationsprofilen des Sekundarbereiches I in bezug auf die Qualifikationsanforderungen der Ausbildungsbetriebe für berufliche Bildungsgänge. (Projektbericht) Köln: (Manuskriptdruck 433 S.) 1981. (c)

- JÄGER, R.S./JUNDT, E.: Mannheimer Rechtschreib-Test (M-R-T). Göttingen 1973, <sup>2</sup>1981.
- JAHN, M.: Studienanfänger 1975 und 1981. In: Mitteilungen des Hochschulverbandes 15 (1982), S. 163–164.
- Kammerreport*: In Deutsch und Rechnen mangelhaft. In: Unsere Wirtschaft (IHK Düsseldorf) 52 (1981), S. 153–154.
- KLAUER, K.J.: Kriteriumsorientierte Tests. In: FEGER, H./BREDEKAMP, J. (Hrsg.): Messen und Testen. Band B, I, 3 der Enzyklopädie der Psychologie. Göttingen 1983.
- KLEIN-BRALEY, C./LÜCK, H.E.: The development of the Duisburg English Language Test for Advanced Students (DELTA). In: GROTHJAHN, R./HOPKINS, E.: Empirical Research on Language Teaching and Language Acquisition. Bochum 1980. S. 85–126.
- KLEIN-BRALEY, C.: Placement test scores as indicators of achievement in English at the Abitur level. In: AKS-Rundbrief (1981), Nr. 1, S. 33–36.
- KLEIN-BRALEY, C.: Report Placement Test Sommersemester 1983. Uni Duisburg, vervielfältigt, 1983.
- KLEIN-BRALEY, C.: Delta Placement Test Report WS 1983/84. Uni Duisburg, vervielfältigt, 1984.
- KRAUSE, F./REINERS-LOGOTHETIDOU, A.: Der bundesweite Studieneingangstest Physik 1978. In: Physikalische Blätter 35 (1979), S. 495–510.
- KRAUSE, F./REINERS-LOGOTHETIDOU, A.: Kenntnisse und Fähigkeiten naturwissenschaftlich orientierter Studienanfänger in Physik und Mathematik. Uni Bonn 1981. (a)
- KRAUSE, F./REINERS-LOGOTHETIDOU, A.: Kenntnisse und Fähigkeiten naturwissenschaftlich orientierter Studienanfänger in Physik und Mathematik. Universität Bonn 1981 (Manuskriptdruck). (b)
- KULTUSMINISTERIUM RHEINLAND-PFALZ (KM): Empfehlungen zur Rechtschreiberziehung. Mainz 1984 (ohne Paginierung).
- KURZ, G. (Hrsg.): Report 18: Dokumentation: Das Eingangswissen in „Mathematik“ und „Physik“ von Studienanfängern an Fachhochschulen technischer Richtung. Furtwangen und Karlsruhe 1981.
- KURZ, G. (Hrsg.): Mathematik für Nichtmathematiker: Hürde am Studienbeginn. Erfahrungen – Modelle – Konzepte. (Hochschuldidaktische Materialien 86). Hamburg 1982.
- KURZ, G. (Hrsg.): Orientierungsverfahren für Studienanfänger im Fach Mathematik. Esslingen: Fachhochschule für Technik 1983 (Manuskriptdruck).
- KURZ, G./BRENNE, H.J., u. a.: Das Eingangswissen in Mathematik und Physik von Studienanfängern an Fachhochschulen technischer Richtung. In: Die neue Hochschule 22 (1981), S. 12–16.
- LEHMBERG, H./LOCHHAAS, H./PAGNIA, H.: Physikkenntnisse angehender Ingenieurstudenten. In: Der mathematische und naturwissenschaftliche Unterricht 35 (1982) S. 305–307.
- LEPPIG, M.: Mathematikleistungen von Studienbewerbern 1976. In: Zentralblatt für Didaktik der Mathematik 10 (1978), S. 101–105.
- LESCHINSKY, A.: Die Steuerung der Schulentwicklung seit den zwanziger Jahren und ihre Schwierigkeiten. In: Vierteljahreshefte für Zeitgeschichte 30 (1982), S. 27–81.
- METZGER, R.: Berufsanfänger: Mängel in den Grundlagen. In: Wirtschaft und Berufserziehung 32 (1980), S. 325–330.
- MÖNIKE, W.: Schüler an der Schwelle zur Berufsausbildung sind besser als ihr Ruf. In: Gewerkschaftliche Bildungspolitik 32 (1981), S. 288–295.
- NÄGERL, H./BECKER, H./HARTEN, H.-U./SCHULTE, H.-D./ZERBST, J.: Über die Schwierigkeiten der Studienanfänger in Medizin im Umgang mit der Mathematik. In: Didaktik der Mathematik 2 (1973), S. 143–157.

- NÄGERL, H./HARTEN, H.-U.: Über die Entwicklung der mathematischen Vorkenntnisse von Studienanfängern der Medizin – oder: Wirkungen der Oberstufenreform. In: *Der Mathematische und Naturwissenschaftliche Unterricht* 33 (1980), S. 175–177.
- PAULSEN, F.: Geschichte des gelehrten Unterrichts. Leipzig <sup>2</sup>1896/97.
- QUEISNER, R.: Der schulische Leistungsstand unserer Lehranfänger. In: *Pfälzisches Industrie- und Handelsblatt* 42 (1967), S. 281–284.
- REIß, M./STEINER, H.-G.: Mathematikkenntnisse, Leistungsmessung, Studierfähigkeit. (IDM9) Köln 1984.
- RUNNE, E.: Wichtige Grundkenntnisse weiter mangelhaft. In: *Wirtschaft und Berufs-Erziehung* 34 (1982), S. 86–89.
- SCHMIDT, J.: Mathematik-Vorkenntnisse als hochschuldidaktisches Problem. In: KURZ, G. (Hrsg.): *Mathematik für Nichtmathematiker: Hürde am Studienbeginn, Erfahrungen – Modelle – Konzepte*. (Hochschuldidaktische Materialien 86.) Hamburg 1982, S. 62–84.
- SCHNOTZ, W.: Zur Schätzung von Aufgabenschwierigkeiten durch Lehrer. In: *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* 3 (1971), S. 106–120.
- SCHÖNEBERG, E.: Erschreckendes Leistungsniveau bei mathematischen Grundkenntnissen. In: *Der berufliche Bildungsweg in Nordrhein-Westfalen* 22 (1981), S. 213–219.
- SCHOENES, F. J.: Eingangstest Physik bei Studienanfängern technischer Fachhochschulstudiengänge während der letzten 3 Jahre. In: SCHARMANN, A., u. a. (Hrsg.): *Vorträge der Frühjahrstagung 1983 Gießen*. Deutsche Physikalische Gesellschaft 1983, S. 610–615.
- SCHULTZ, T., u. a.: *Theorie, Experiment und Versuchsplanung in der Psychologie*. Stuttgart 1981.
- SCHWARZE, J.: Über Mathematikkenntnisse von Studienanfängern. In: *DGOR-Bulletin* Nr. 27. (1985), S. 10–11.
- SCHWIBBE, G./SCHWIBBE, M.: Validierungsuntersuchung zum Duisburg English Language Test for Advanced Students. In: GROTHJAHN, R., u. a. (Hrsg.): *Taking their measure: the validity of language tests*. Bochum 1986 (im Druck).
- STEINER, H.-G.: Mathematisch-naturwissenschaftliche Bildung. In: REIß, M./STEINER, H.-G. (Hrsg.): *Mathematikkenntnisse, Leistungsmessung, Studierfähigkeit*. Köln 1984, S. 5–56.
- STEMME, H.: Zum Leistungsstand der Berufsanfänger. In: *Archiv für Berufsbildung* 6 (1954), S. 94–96.
- Symposion „Verbesserung der Rechtschreibleistungen“*. Bad Kreuznach: Arbeitsstelle für Lehrplanentwicklung und -koordination des Landes Rheinland-Pfalz 1984.
- WALTHER, R. H.: Die Rechen- und Rechtschreibkenntnisse der Volksschulentlassenen. In: *Jugend und Beruf* 3 (1929), S. 438–443.
- WIRTSCHAFTSVEREINIGUNG EISEN- UND STAHLINDUSTRIE: Abnehmende Schulkenntnisse gefährden die Berufsausbildung. Düsseldorf: Januar 1982 (vervielfältigt).
- WITTMANN, B.: In der „Ohrghelmusik“ klingen Mißtöne. In: *Neue Deutsche Schule* (1967), S. 170–176.
- WITTMANN, B.: Junger Wein in alten Schläuchen – Bereiten unsere Schulen noch ausreichend auf den Beruf vor? In: *Neue Deutsche Schule* 33 (1981), S. 26–27.
- Ohne Autor: Schwaches Fundament erschwert Berufsausbildung. In: *Kurhessische Wirtschaft* 23 (1976), S. 653–654. Identisch mit dem Artikel in: *Wirtschaft und Berufs-Erziehung* 29 (1977), H. 1, S. 21–23.
- Ohne Autor: Ergebnisse des Rechtschreib- und Rechentests 1979. In: *Kurhessische Wirtschaft* (1980), H. 2, S. 54–55.
- Ohne Autor: In Deutsch und Rechnen mangelhaft. In: *Unsere Wirtschaft* 52 (1981), S. 153–154.



Ohne Autor: Rechtschreiben und Rechnen weiter unbefriedigend. In: Kurhessische Wirtschaft (1981), H. 1, S. 10–12. Identisch mit dem Artikel in Wirtschaft und Berufs-Erziehung 33 (1981), S. 141–143.

Ohne Autor: Über die Qualifikation von Ausbildungsstellen – Bewerber. In: Betriebliche Ausbildungspraxis 28 (1982), S. 929–931.

### *Abstract*

*The Discussion about the Achievement of Students Entering Vocational Training or Universities – A Critical Review of Research and Statements in the Federal Republic of Germany.*

After having left school after 8th grade (nowadays mostly after 9th or 10th grade), German students start their vocational training in factories, shops etc. accompanied by part time vocational instruction in schools.

Getting accepted by a university requires as a rule 13 years of formal instruction in schools.

A historial review shows that during the last centuries each generation had claimed that schools were achieving less.

A critical examination of this assertion leads to the conclusion that no systematic programs do exist in Germany to support evidence of changes in school achievement. From a methodological standpoint, the few studies available are unsatisfactory and cannot validate either increases or decreases in school achievement.

### *Anschrift des Autors:*

Prof. Dr. Karlheinz Ingenkamp, Trifelsstraße 42, 6741 Leinsweiler